DOCUMENT RESUME

ED 377 210 TM 022 368

AUTHOR Paulson, F. Leon; Paulson, Pearl R.
TITLE A Guide for Judging Portfolios.
PUB DATE May 94
NOTE 16p.
AVAILABLE FROM Measurement and Experimental Research Program, Multnomah Education Service District, 11611 N. E. Ainsworth Circle, Portland, OR 97220.
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary School Students; *Evaluation Methods; Grade 2; Mathematics; Pilot Projects; *Portfolios (Background Materials); Primary Education; *Scoring; Student Evaluation; Test Construction
IDENTIFIERS *Cognitive Model for Assessing Portfolios; *Performance Based Evaluation; Stakeholders

ABSTRACT

A theory of portfolio assembly is presented and used to guide analysis of portfolios. Features of a portfolio evaluation approach known as the Cognitive Model for Assessing Portfolios (CMAP) are outlined. This model assumes that portfolios begin as relatively simple collections and gradually grow into complex, articulate stories of knowing. It is a purposeful, integrated collection of student work that shows student effort, progress, and achievement in one or more areas. CMAP is an organizing theory that serves as a lens through which one can view and think about any portfolio. It places the processes associated with portfolios into three major categories of concern: stakeholder, process, and history. A four-stage rubric is presented for evaluating portfolios, corresponding to the following stages of portfolio growth (1) the off-track portfolio; (2) the emerging portfolio; (3) the on-track portfolio; and (4) the outstanding portfolio. A pilot study involving 42 second-grade mathematics portfolios and 3 raters illustrates application of the model and rubric. Three figures and six tables illustrate the discussion. Contains 26 references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 377 210

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
F. L. PAULSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

A Guide for Judging Portfolios

F. Leon Paulson
Multnomah ESD

Pearl R. Paulson
Beaverton SD

May 1994

SEVERAL STUDIES THAT USE portfolios to assess learning look at learning outcomes reflected in portfolios without looking at portfolios themselves (e.g., Gearhart, Herman, Baker & Whittaker, 1992; Gentile, 1992; Vermont 1991-a,b). The scoring rubrics are modifications of those used in existing performance assessment that assess relatively traditional outcomes such as how well students write. But to many, the advantage of using portfolios is the process portfolios bring to the classroom. The portfolio process is highly interactive —between student and teacher (Frazier & Paulson, 1992), student and other students (Kauffman & Short, 1993), and within the individual student through self–evaluation (Paulson and Paulson, in press: Rief, 1990). This aspect of portfolio is not being assessed in these studies. In fact, in the Gearhart et al. study, the authors note that reviewers expressed surprise that student self–reflections and other process materials had been removed from the portfolios prior to review.

An evaluator must be clear about what he or she is looking at when doing a portfolio study. Portfolios vary from folders of student work gathering dust in the corner to dynamic collections of student work that are an integral and necessary part of the daily classroom activity. Little is gained by lumping them together for the purpose of analysis. This paper presents a judgment rubric designed to assess the quality of the portfolios themselves in a way that provides a context for interpreting the influence of portfolios.

It is possible to think of a portfolio as merely a bunch of "stuff" collected by teachers or students. But most teachers would be dissatisfied with this definition of portfolio. It is not stuff — it is a rich vein of information on the student just waiting to be mined. During a workshop, hand a group of teachers a portfolio and they hungrily paw through it, discovering an enormous amount about the student, the student's learning, and the educational program from which the portfolio came.

But give the portfolio to different groups of teachers and you will discover that they often learn very different things. Often they agree, but frequently they disagree. We are not particularly troubled by the fact that people reviewing portfolios disagree about what they see. To us, the portfolio may be a means for interpreting the ill defined, an attempt to get

Copies of this document are available through the Measurement and Experimental Research program of the, Multnomah Education Service District, 11611 N. E. Ainsworth Cir. Portland, OR 97220.

BEST COPY AVAILABLE

TM022368

interpreting the ill defined, an attempt to get experience to 'hold still' long enough so that we can get a rough idea of what is going on (Schulman, 1992). What does concern us, however, are conclusions that are made without documentation and that ignore the frame of reference adopted by those making the judgments.

This paper adopts a theory of the way portfolios are assembled and uses that theory to guide the analysis of portfolios. The theory presents a way of thinking about portfolios that makes 'the items found in them interpretable and meaningful.

This paper presents some of the features of a portfolio evaluation approach known as the Cognitive Model for Assessing Portfolios or CMAP (See F. L. Paulson & Paulson, 1990; Paulson, Paulson & Frazier, in press; P. R. Paulson & Paulson, 1991). It infers cognitive properties associated with portfolio development. We assume that portfolios begin as relatively simple collections of things that gradually grow into complex, articulate stories of knowing.

Portfolios often begin when teachers ask students to collect work in folders. The process is usually teacher directed and may involve little active student participation. Over time, the process of collecting and evaluating can become more and more involved as the student becomes absorbed in the process. Our conceptual model describes this process and offers a rubric that can be used when making judgments about individual portfolios. Many of its concepts are from cognitive psychology — that part of psychology that deals with the way people think, learn, and create knowledge. We hope our model will help illuminate some of portfolio assessment's murky areas.

Here is what we mean by portfolio: *A portfolio addresses the question "who am I" and tells a coherent story of the student as learner. It is a purposeful, integrated collection of student work that shows student effort, progress, or achievement in one or more areas. The collection includes evidence of student self–reflection and student participation in setting the focus, establishing the standards, selecting contents, and judging merit. A portfolio tells the student's own story of what is learned and why it is important.*[1]

This somewhat complex definition can be paraphrased: *A portfolio tells a story, it is a story of learning. A portfolio contains anything that helps tell that story.*

The Cognitive Model for Assessing Portfolios (CMAP)

CMAP[2] is a framework for thinking about and evaluating a portfolio in a coherent fashion. It is an organizer, a theory around which activities and portfolio contents can be understood. CMAP does not dictate a specific format for portfolios. Rather, it is a lens through which one can view and think about any portfolio. CMAP is an evaluation model that uses the concepts presented in the rubric and the process model presented in this paper.

CMAP places the processes associated with portfolios into three major categories of concern. We think of these three dimensionsas a kind of topographical map that reflects the portfolio itself and the context in which it exists. The three dimensions involve (1) the people, (2) the processes, and (3) the record over time.

- *Stakeholder*

  Portfolios are created by students working together with teachers, often under the watchful eye of others. The stakeholder dimension identifies viewpoints of individuals or groups who have an interest in the

[1] This definition is an expanded and refined version of the widely quoted definition developed under the auspices of the Northwest Evaluation Association (See Paulson, Paulson, & Meyer, 1991).

[2] Patterned after Robert Stake's (1967) program assessment model, we introduced CMAP in a theoretical paper *How do portfolios measure up: The Cognitive Model for Assessing Portfolios* (Paulson & Paulson, 1990).

portfolio. Stakeholders are more than an audience— they are participants (see Guba & Lincoln, 1989). The student is central or *primary* stakeholder. Secondary stakeholders may be teachers, parents, assessment specialists, and others. They may play a positive, supporting role or they may have a negative impact on the process.

- *Process*

Students and teachers perform many kinds of tasks when putting together a portfolio. The process dimension describes the activities involved in building a purposeful, interrelated collection of student work. They include stating the purpose or rationale for the portfolio, deciding on specific issues or themes to be treated, establishing standards to be used in judging the portfolio, collecting the items that comprise the actual portfolio, and making the entire collection meaningful.

- *History*

While some portfolios deal with end states or outcomes, i.e., very best work, many instructional portfolios are more process–oriented with an interest in how that best work evolved. The historical dimension looks at changes over time. It looks at conditions at the outset (antecedent conditions), what activities occur during the time the portfolio is assembled (transactions), and what happens as a result (outcomes). Any or all processes on the stakeholder and activity dimensions have a historical perspective.

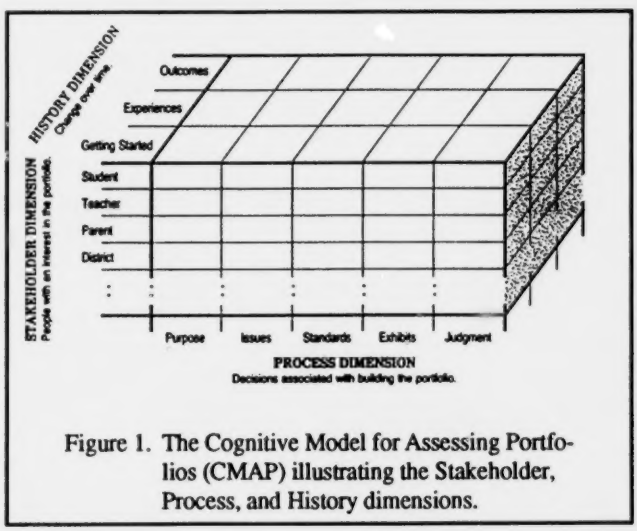

Figure 1. The Cognitive Model for Assessing Portfolios (CMAP) illustrating the Stakeholder, Process, and History dimensions.

Figure 1 is 3-dimensional representation of CMAP which gives a comprehensive view of the portfolio. Each dimension functions in concert with the remaining dimensions. When one considers, for example, the reason for creating a portfolio, CMAP reminds us that each stakeholder may hold a somewhat different purpose and that those purposes may change over time. CMAP as presented here is abstract and conceptual. For an example of how CMAP might look in actual practice, see *Sarah's Portfolio* by Paulson, Paulson, & Frazier (in press) which decribes an actual portfolio assembled by a student.

On Using The Rubric

The rubric which appears on the next page is a guide for making judgments about portfolios using some assumptions about cognitive processes underlying portfolio development. The processes involved in creating a portfolio are far more important than the portfolio's contents. However, cognitive processes are not observable — they must be inferred from examining the products found in the portfolio. The rubric helps you infer underlying processes by guiding your examination of portfolio contents. Thus the rubric provides a frame of reference for your judgments, not an infallible guide.

Let us illustrate. Our definition refers to a portfolio as a story. A story is an organized presentation of some sort, a communication between the portfolio builder and the reviewer. If you review a portfolio and judge it to tell a coherent story, you can infer that the portfolio builder went through a process associated with creating an organization and fitting the parts

BEST COPY AVAILABLE

Table 1
The Four Stages of Portfolio Growth
A Summary of the Rubric for Judging Portfolios

An Off–Track Portfolio

An off–track portfolio is simply a container of student work or assessments, without an attempt on the part of the learner to provide organization. There is no attempt by the learner to make a coherent statement about what learning has taken place. The child's understanding of the task is minimal — the portfolio is about "collecting what the teacher asks for." For the student, the portfolio was built by following instructions. Self–reflective statements if present add little to clarify organization or explain learning.

An Emerging Portfolio

In an emerging portfolio there is a sense of intentionality controlling some of the student's choices. Students may not be able to verbalize the reasons, even as they reflect on their choices, but the reviewer may be able to recognize a relationship between some exhibits or infer the reasons. Or, there may be evidence that the student had some insight into the teacher's purposes. While evidence of self–reflection adds information to the presentation, at this point in the development of the portfolio there is insufficient information or organization to characterize the portfolio as either a story of learning or a portrait of the learner.

An On–Track Portfolio

An on–track portfolio is in the process of becoming a story of the student as an independent learner. There are relationships between one part of the portfolio and another. There is evidence of student ownership. The learner has a personal investment in selecting and explaining the content. It is possible to distinguish other stakeholders' goals from the student's or to recognize instances when they overlap. The portfolio may be created for others to assess, but there is also evidence of self–assessment. The student's voice is always audible.

An Outstanding Portfolio

An outstanding portfolio is a coherent story of the student as a reflective learner where all the parts of the portfolio bear a clear relationship to each other and to a central purpose. There is an awareness of the perspectives of other stakeholders, and the student's self–assessment has been enhanced by this knowledge. A reviewer can look at the portfolio and easily understand how the judgments about the learner came to be made and the degree to which different stakeholders would agree. When reviewing the portfolio, outsiders get the feeling they really know the person whose achievement is depicted there, and have a fair understanding of how the learning came about.

---

together. There are many indicators that a portfolio is organized, e.g., a table of contents, an introductory letter, section dividers. However, the presence of these indicators does not guarantee that the portfolio is organized any more than their absence indicates an absence of organization. A creative student can invent many ways to organize something, an uninspired student following directions can mindlessly produce all of the indicators without ever reflecting on or understanding their purpose. In one case, a portfolio with none of the indicators could be judged outstanding, in the other, a portfolio with all of the indicators could be judged off–track. The final judgment requires what Eisner (1991) calls *connoisseurship*. As a judge's understanding of the portfolio and the processes associated with portfolios increases, you become better able

Table 2
Four Kinds of Self–Reflection
Found in Portfolios

The four varieties of self–reflection occurs in the context of specific kinds of portfolio activities that occur in the classroom. They are found across the entire K–12 spectrum. For a more detailed description, see Paulson and Paulson (in press; 1992-b).

Documentation:

Students talk (and write) about why they selected a specific item for the portfolio. Typical comments are

> *I like it.*
>
> *I got a good grade.*
>
> *It is more me than any of my other writing.*

Comparison

Students compare two or three specific items in the portfolio and talk about similarities and differences. It is more interpretive than *Documentation.* This first grader comparing two writing samples of his writing

> *At the beginning of the year. I havit been yoosng periods and I am now. At the beginning of the year. I havit been yoosng sentence and I am now. At the beginning of the year. I havit been been yoosng elaboration and I am now the End.*

Integration :

Students talk about the entire portfolio and what it tells in an overall sense. Here is an example from a learning disabled fourth grader

> *I have really improved in my writing....I can spell bigger and better words now....I also help people edit at their stories. When I was younger I jumped around and told lots of details but now I can stay on order and tell less....*

Presentation:

Students talk about their portfolios from the perspective of others who are reviewing their portfolio. Student–led conferences are among the activities that are effective in involving students in this kind of self–reflection.

> *Dear fifth grade teacher*
>
> *I think I an a fast learning student. I can take a boring report and turn it into a great experience for everyone. I do well in groups, spelling, math, reading, science, and especially writing...I think of myself as a real adventure. I hope you will to.*

---

to identify the quality of a portfolio much as a wine connoisseur learns to identify a good vintage.

Finally, we would like to discuss the question of whether to use a holistic or analytic rubric when judging portfolios. This is often an issue which is hotly debated. Our view is that you select the type of judgement which is appropriate to the item being judged. We have chosen a holistic rubric because we review portfolios as holistic entities that are integrative rather than analytic in nature. While we have observed individual "traits" (e.g. the presence self–reflective statements) in portfolios, we have found them to vary widely across individual portfolios. Thus, we recommend that an evaluator interested in analyzing specific portfolio characteristics do so separately from judging the overall ability of the portfolio to tell a coherent story. An outstanding portfolio is, in its totality, a self–reflection. In the pilot study described below, we use the holistic rating of the portfolios (the ability to

communicate a story of knowing) to provide a context in which to interpret our more analytic analysis of the portfolio contents.

Portfolios and Self–Reflection

An important aspect of the portfolio is that the portfolio itself becomes medium of learning, an opportunity for the student to engage in self–reflection and to learn self–assessment. From our point of view, self–reflection is integral to the portfolio. Accordingly, we have incorporated self–reflection into the overall judgment system. In our work with portfolio projects, we have identified four kinds of self–reflection, each important in its own right, but each making a special kind of contribution to cognitive processes associated with portfolio growth. Table 2 presents a brief description of these four kinds of self–assessment (See Paulson & Paulson, in press; 1992–b).

Interpreting self–reflection in portfolios is tricky business. Students can produce very impressive self–reflective statements without doing much actual self–reflecting. This can happen, for example, when teachers in a legitimate attempt to teach self–reflection skills provide students with models, prompts, and checklists that themselves become substitutes for thinking and problem solving (See Paulson & Paulson, in press; 1992–a,b for a discussion of this issue). The challenge is to look for evidence that the student is engaging in original thinking rather than following a formula. *This requires the judge to look at the self–reflective statement within the context of the complete portfolio*. Therefore, while this judgment rubric is designed specifically to be sensitive to student self–reflection, it does not make judgments about self–reflection as an independent component of the portfolio.

The Four Stages of Portfolio Growth

A portfolio does not appear fully developed in the classroom. It grows. For the purpose of assessment, we have divided that growth into four stages. The first stage is the familiar "folder of material." We call this an *off–track* portfolio because it does little to encourage the use of the portfolio in learning. The second stage occurs when the folder begins to show the first substantial signs of becoming a true portfolio. We call this the *emerging* portfolio. In the emerging portfolio we begin to see the portfolio being used as an environment associated with instruction and assessment. We call the third stage the *on–track* portfolio. In an on–track portfolio, the learner shows evidence of self–direction and self–assessment. The portfolio is beginning to communicate a definite story of learning. The final stage, *outstanding*, reveals a fully developed, mature portfolio.

We now turn to a description of the four stages (a summary appears in Table 1).

Stage 1: The Off–Track Portfolio

An off–track portfolio is simply a container of student work or assessments, without an attempt on the part of the learner to provide organization. There is no attempt by the learner to make a coherent statement about what learning has taken place. The child's understanding of the task is minimal – the portfolio is about "collecting what the teacher asks for." For the student, the portfolio was built by following instructions. Self–reflective statements if present add little to clarify organization or explain learning.

The off–track portfolio is the familiar folder of unorganized material found in many classrooms. The child's understanding of the task is minimal — the portfolio is about "collecting stuff that the teacher wants." The teacher may be following guidelines from the central office or materials from a publisher. Whatever the case, someone other than the student, some secondary stakeholder, is calling the shots. To the child, building a portfolio is done by following instructions.

Since our definition of portfolio emphasizes process over product, it is important to consider how the process gets started and under what conditions it stops (CMAP's history dimension). The off–track portfolio is initiated by the teacher and is guided by goals set by the teacher. The teacher also decides when the portfolio is finished, a process that might be paraphrased, "I've

[teacher] got all I need, you [student] can stop collecting." Using the language of CMAP, the student's involvement is limited to putting things into the portfolio, an activity limited to the contents cell of the *process* dimension. The student shows little evidence of a sense of purpose and with little understanding of the issues or that entire collection as a presentation of self which would be indicated if the student were participating in processes associatged with all cells on the process *dimension.*

Stage 2: The Emerging Portfolio.

In an emerging portfolio there is a sense of intentionality controlling some of the student's choices. Students may not be able to verbalize the reasons, even as they reflect on their choices, but the reviewer may be able to recognize a relationship between some exhibits or infer the reasons. Or, there may be evidence that the student had some insight into the teacher's purposes. While evidence of self–reflection adds information to the presentation, at this point in the development of the portfolio there is insufficient information or organization to characterize the portfolio as either a story of learning or a portrait of the learner.

The emerging portfolio represents more complex mental processes that result when the student becomes a more active participant. The child is beginning to view the portfolio as something more than a class assignment. A concept of "portfolio" begins to take shape as the student begins to self–monitor and performing executive functions associated with building the portfolio.[3] While the portfolio activity itself was probably initiated by the teacher and proceeds under the general guidance of the teacher's concept of portfolio, the student is gradually developing a task representation of what it means to "do a portfolio." In fact, the student may become quite independent at this stage, but that independence remains focused on item selection. Doing a portfolio involves reinventing the wheel — the benefit to the student is that it becomes the student's wheel.

A very important process begins to appear at a rudimentary level in the emerging portfolio, that of explaining to self and others the reasons for selecting items (*Documentation*). The explanations are probably simply telling with little attempt to interpret. The role of the teacher in this process is particularly important. Vigotsky (1962) suggests that all higher order skills first appear in interactions between the child and other. As a result of the interactions, the child gradually internalizes these higher order skills. Vigotsky's insight may hold the key to understanding the role of the portfolio in instruction and assessment, that of providing the opportunity for dialogue which will lead gradually to the development of higher order thinking skills.

At this stage, the child begins to develop a kind of mental model of the portfolio. This model provides the basis for selecting material for the portfolio. At first, this mental picture may guide specific decisions about what to place into the portfolio. This process can be expected to gradually become more complex both as the child grows older and as the child becomes more experienced with portfolios.

The point at which work on the emerging portfolio stops is probably determined by the teacher although the student may have some input. Unlike the off–track portfolio, work stops because the student or the teacher compares the actual portfolio with the conceptual model of the portfolio and find the two congruent. External influences such as the end of the school year might interrupt work or hasten the schedule, but not define "a completed portfolio."

It is important to note that many of the activities found in the off–track portfolio are found in the emerging portfolio as well. Collecting continues but with elements that make it a more com-

[3] *Executive functions* refer to cognitive or thinking abilities that integrate other thinking abilities. For example, a student may be able to write a passage, and also be able to edit a passage in order to improve it, but may not know to switch from one mode (writing) to the another (editing) at the appropriate time without being told to do so by a teacher (Bereiter & Scardamalia, 1987). Portfolio assessment by encouraging self–monitoring helps students learn at the executive (when to write, when to edit) as well as at the skill level (how to write and how to edit).

plex process. One of the major differences between this and the off–track is that the student begins to assume some of the more complex, decision making functions associated with planning and taking charge of the portfolio.

When a portfolio begins to tell an overall story and the student becomes involved as an active participant, integrative activities begin to appear. They begin to appear when the student makes *Comparisons* (e.g., reviewing pieces done early and late in the year and commenting on how much has been learned). This is the first appearance of one of the most important and central processes associated with higher stages of portfolio activity. Many emerging portfolios contain *Integrative* self–reflections, early attempts to talk about the portfolio as a whole.

Explained in CMAP terms, in the emerging portfolio, more of the activity dimension is evident. Some of the elements of story can be inferred. For at least some of the contents, there seems to be a uniting purpose, and certain exhibits seem to pertain to the same issue (interest, goal, desired outcome). However, the other dimensions of the model are missing from this stage. The portfolio gives no importance to whether the purpose, issues, or reviews are the student's own or belong to some other stakeholder. Similarly, the historical dimension has limited relevance although it may be possible for a reviewer to infer how the quality of the student's work changes when sampled at different times.

Stage 3: The On–Track Portfolio

An on–track portfolio is in the process of becoming a story of the student as an independent learner. There are relationships between one part of the portfolio and another. There is evidence of student ownership. The learner has a personal investment in selecting and explaining the content. It is possible to distinguish other stakeholders' goals from the student's or to recognize instances when they overlap. The portfolio may be created for others to assess, but there is also evidence of self–assessment. The student's voice is always audible.

An on–track portfolio is simply an embryonic full–fledged portfolio. Thus, the discussion of the cognitive processes associated with the outstanding portfolio also applies to the on–track portfolio as well. The major difference is that while the on–track has several attributes of an outstanding portfolio, the truly outstanding portfolio has most of them.

In terms of the CMAP model, an on–track portfolio provides not only a sense of story, but also the student's voice telling that story. At this stage, the stakeholder dimension first becomes pertinent: We can follow the activity dimension from the perspective of more than one stakeholder. Most particularly, we can distinguish that at least some of the activities are the student's. There is evidence of student ownership of purpose, issues, choices of content or review. Often, one notes shifts in the relative influence of different stakeholders at various points on the activity dimension. For example, some content may be chosen according to the student's purposes, others aligned to specific teacher issues. However, in the on–track portfolio, the historical dimension is not yet fully developed. While an inspection of content gathered at different times may reveal progress, or a review bring about some redirection of purpose or clarification of issues, it is not clear whether students are aware that their learning has been illustrated from information in such reviews, or from the input of other stakeholders.

Stage 4: The Outstanding Portfolio

An outstanding portfolio is a coherent story of the student as a reflective learner where all the parts of the portfolio bear a clear relationship to each other and to a central purpose. There is an awareness of the perspectives of other stakeholders, and the student's self–assessment has been enhanced by this knowledge. A reviewer can look at the portfolio and easily understand how the judgments about the learner came to be made and the degree to which different stakeholders would agree. When reviewing the portfolio, outsiders get the feeling they really know the person whose achievement is depicted there, and have a fair understanding of how the learning came about.

Both on–track and outstanding portfolios represent the emergence of a fully documented student presentation. The on–track portfolio has some of the elements about to be described, and outstanding portfolio has many of them.

The student has an overall purpose of the portfolio that guides the overall enterprise, yet the purpose itself is under review and changing. The model suggests that one factor of interest is the relationship between the student's concept and representation and that of the teacher. Similarly, the student selects items that address issues and satisfy standards, the student's own and quite possibly those of other stakeholders (e.g., teachers, parents).

The self–reflective writings that students attach to exhibits move toward knowledge processing[4] which includes interpretation and documentation. As the student becomes more interpretive, the self–reflections begins to influence the way things are selected and judged. Through self–reflection, the student has new ways to look at the issues and standards used in the decision making process itself. The student assumes more ownership and investment in the portfolio. As students revises the standards and procedures associated with building the portfolio, they are more likely to make the portfolio "their own."

At this stage, the *Integration* and possibly *Presentation* forms of self–reflection will show that students view the entire portfolio as a cohesive document rather than as a collection of individual pieces. They also are able to take the interests of other stakeholders into account and become concerned with providing a history of their learning. They sort and organize.[5] Bruner (1963) argues that understanding the underlying structure of knowledge (i.e., how things relate to one another) is more important than knowing content. The on–track and outstanding portfolios begin to tell about how knowledge is organized, integrated, connected, and used rather than just how much of it a student has.

The question of deciding when the portfolio is complete is no longer a simple "yes/no" decision. Rather, it is a matter of making complex judgments about the way the idealized portfolio compares to the real one. The portfolio can be considered unfinished in a variety of different ways, each with different implications. For example, the student may be pleased with the story the portfolio tells but feels that the presentation needs improvement. Thus, the student may make modifications of a stylistic nature — to achieve better "packaging" of the story. The process of judging a portfolio may involve writing an extended self–reflective statement about the overall portfolio (*Integration, Presentation*). The processes associated with the student documenting the judgment involve self–reflection of a very different type from that observed in item selection (*Documentation*) or in comparisons made on a single dimension (*Comparison*). Students are asked to look at many different aspects of the portfolio and make overall, general judgments.

The owner of a truly outstanding portfolio may never achieve closure in the sense of final completion. This might be paraphrased as "It's not finished, but it's going very well..." (In a true story of learning, the paraphrase might read "its not are finished because I'm not finished.")

Pilot Study

Let us look at some preliminary results using this rubric to judge actual portfolios. These data pertain to the accuracy (validity) and trustworthiness (reliability) of information generated using

[4] Bereiter and Scardamalia (1987) identify two kinds of cognitive processes associated with writing — knowledge telling in which the writer simply puts words on paper and knowledge processing in which the writer uses the writing process itself as a way to think through issues.

[5] It is possible that encouraging children to engage in tasks like constructing tables of contents and the like will have little affect until the child begins processing at the integrative level. The same may be true for integrating some of the external concerns associated with the stakeholder dimension such as materials prepared to be shared at student–led conferences.

the rubric. We recommend approaching the questions of reliability and validity with caution. Portfolios are not academic achievement tests and we suspect that the rules that govern the construction and interpretation of traditional tests may require rethinking before they can be applied to portfolios (see Harman, 1992; Moss, 1994; Paulson & Paulson, 1991; and Wolf, Bixby, Glenn & Gardner, 1991 for cautions on applying traditional psychometric techniques to alternative forms of assessment).

The question of reliability is singularly difficult to address in portfolios. One can readily collect data on the level of agreement among judges, but within the context of the portfolio it is not immediately apparent how to interpret them. Earlier, we talked about portfolios as a means of interpreting the ill–defined. They are complex, integrative things, and different stakeholders can reasonably be expected to bring different but completely valid perspectives to the task of judging them. Thus, disagreement may be useful information to be preserved rather than measurement error to be eradicated. Elsewhere (Paulson & Paulson, 1991) we refer to the Siskel and Ebert approach to making judgments (named for the film critics on television whose trademark is colorful debates regarding the quality of the films they review) in which we argue that by encouraging disagreement, one can actually add information, clarification, and understanding to the assessment process. Other ways typically used to study reliability (e.g., equivalent forms, item manipulations) are even more difficult to think about and interpret in the portfolio context. If a portfolio is an integrated whole, how can we legitimately divide it into parts for analysis? A really good portfolio would probably be rendered meaningless by such a manipulation.

The question of validity also poses some interesting dilemmas. What kinds of criteria should we apply to judge their accuracy of judgements made using our rubric? One criterion might be the correlation between portfolio quality as measured on our rubric and traditional measures of achievement. We expect that the ability of a student to put together a quality portfolio is relatively unrelated to academic achievement. We have found high quality portfolios are found in programs in which academic performance may be marginal. Thus, we predict a low correlation between quality ratings of portfolios and achievement test scores.

We used the rubric to judge a sample of 42 second grade math portfolios (14 teachers each provided 3 portfolios). The portfolios were from the first year of implementation of math portfolios in a medium sized district. A team of three teachers (one second grade, two third) from the district studied the rubric, then reviewed the portfolios and placed them into the four judgment categories. In cases where there was disagreement, the three discussed their disagreements and reached consensus on the rating. A second judge later rated the portfolios without referring to the teams ratings.

I (Leon) trained the team and served as the second judge. I did not participate in making the initial ratings, however I did observe the team while they made their ratings. While listening to their discussions, I got the impression that two factors were at work. First, the team seemed determined to use all available categories (off–track, emerging, on–track, and outstanding) even though I had told them during training that there was reason to expect that few if any portfolios would fall into the top (outstanding) category. This is because the project was in its first year and teachers need time to become comfortable with portfolios, and most of the 15 participating teachers had received little training on portfolios. At first, the team judged without using theoutstanding category. The fact that one category (was not being used, however, seemed to bother them and, near the end, they revised some of their judgments, moving 6 portfolios into the outstanding category. They also reconsidered the bottom category, moving some that had been judged off–track to emerging.

Since the portfolios were available for only 5 days, it was impossible to have them independently rated a second time. Therefore, four days

later, I personally rated all 42 portfolios without referring to the judgments made by the group. This is not, of course, an independent judgment which would be required in a formal inter–judge agreement study.[6]

For the purpose of statistical analysis, we translated the descriptive categories into numerical values (off–track = 1...outstanding = 4) in order to calculate agreement between judges. The results appear in Table 3. The two sets of judgments agree on specific categories about half the time. However, when differences greater than one point are ignored (this is called adjacent category agreement, a procedure frequently used in performance assessment studies), agreement is high. The correlation coefficients show that agreement on ranking is also relatively high (.8) suggesting that observers using this rubric rank the portfolios in a similar manner although that there are disagreements with respect to assignment to specific categories. These disagreements, however, are limited to assignment to adjacent categories.

Table 3
Agreement Among Judges
n=42

Complete agreement	52%
Adjacent category (one pt. diff. allowed)	100%
Correlation	.84

The general agreement in rankings coupled with the relative disagreement in assigning to specific categories suggests that the team and the second judge may have judged the portfolios differently in a systematic way. Our hypothesis that the team had changed their frame of reference and moved all ratings upward is supported by the data in Table 4. The second set of judgments seems to have moved a proportion of the judgments "down a notch," but with little changes to the relative rankings.

Table 4
Classification Decisions by Judge
n = 42

Category	Team		Second judge	
Off–Track	16	(38%)	25	(60%)
Emerging	15	(36%)	11	(26%)
On–Track	5	(12%)	6	(14%)
Outstanding	6	(14%)	0	(0%)

There are at least two explanations for the systematic difference in these scores. One is that either the team or the second judge misapplied the rubric. As noted earlier, the team apparently shifted their frame of reference in order to use all available categories. This suggests that the agreement would be increased with better instructions, training, and using benchmark portfolios. On the other hand, these differences may reflect a substantive factor reflecting differences in perspective among judges. The team represented insiders, teachers working in the primary classroom in the district while the second judge was an outsider, an assessment specialist concerned with developing a judgment system with certain formal properties. This suggests that the two sets of data may reflect equally valid stakeholder perspectives that require different interpretations. Differences of this kind can be represented on the stakeholder dimension of the CMAP model.

Let us now turn our attention to data that may provide some information that will help validate the rubric by comparing observed data with our theoretical notions about processes that occur in portfolios. First, our theory suggests that all students are capable of engaging in the kinds of processes described above and that evidence of these processes will emerge in higher quality portfolios.

We had asked teachers to submit portfolios from a range (low, middle, high achieving) of

[6] Do not be perplexed by our inclusion of agreement data following a discussion questioning the way such data are often used in evaluation studies. Such data provide useful information on the frames of reference the judges are using and how data might be interpreted. What we question is the assumption that data from judges who agree is somehow better than data from judges who do not.

students in their classes. While the request was informal and we made no attempt to formally verify their compliance, the impression from looking at individual math worksheets in the portfolios was that the teachers complied. As I watched the judges at work, I got the impression that one could predict the rating given a specific portfolio from knowing how other portfolios from the same teachers had been rated. If the first portfolio from a teacher's class was rated emerging, it was likely that the other two portfolios from the same class would also be rated emerging. Since each teacher submitted portfolios from a range of achievement levels, the quality of portfolios should be unrelated to achievement level. The analysis presented in Table 5 supports this impression. In every instance, all portfolios submitted by the same teacher were judged to be within one point on the rubric suggesting that achievement level of the students may be unrelated to their ability to put together a portfolio.

Table 5
Teachers Receiving Similar Judgments
n = 14

	Complete		Adjacent Category	
Team	9	(64%)	14	(100%)
Second Judge	13	(93%)	14	(100%)

To explore this further, we looked at the portfolios from three teachers who had received a substantial amount of training on portfolios (they had received 36 hours of instruction on portfolios while the others had received only 6 hours).

Table 6
Ratings of Portfolios in Classes Where Teachers Received Additional Training

	All Teacher		Teachers with Add'l Training	
Category				
Off–Track	16	(38%)	0	
Emerging	15	(36%)	0	
On–Track	5	(12%)	3	(33%)
Outstanding	6	(14%)	6	(65%)
Total	42		9	

The results (Table 6) indicate that additional training had a strongly positive impact on the quality of portfolios in the classroom. Again, the achievement level of the students, apparently, had little impact on the quality of the portfolio processes reflected in the judgments.

A second way we examined validity was to look at the kinds of items found in the portfolios. Our theory suggests that as portfolios increase in quality, two things should happen. First, the people assembling the portfolios should become more selective. Second, the kinds of material found in the portfolios should be increasingly likely to reflect individual student involvement and ownership. The pilot study produced one category of data that related to these hypotheses. We asked the team to count all items in the portfolios (e.g., including work samples, letters to the reader). If the process of choosing items for the portfolio becomes increasingly selective (that is, the student becomes better and better at making decisions about what to put into and take out of a portfolio), then the number of items selected will be likely decrease as portfolios improve in quality. Second, we then asked them to count the number of work samples which fell into two categories: Worksheets (teacher or commercially prepared items that imply less personal involvement by the child); or child–produced (materials entirely in the child's own hand implying more personal involvement). Our theory predicts that the ratio of child–produced items to worksheet items will increase with portfolio quality.

Figure 2 presents the results. It appears that as students and teachers work together to improve the quality of the portfolios, the first thing that happens is that they put more items into the portfolios. But as the portfolios move into the outstanding category, this trend is sharply reversed as would be expected if they are becoming more selective. In fact, one outstanding portfolio contained only 11 items while one off–track portfolio had 40. Some good stories, apparently, can be told in a few words. Interestingly, the number of student–produced items remained

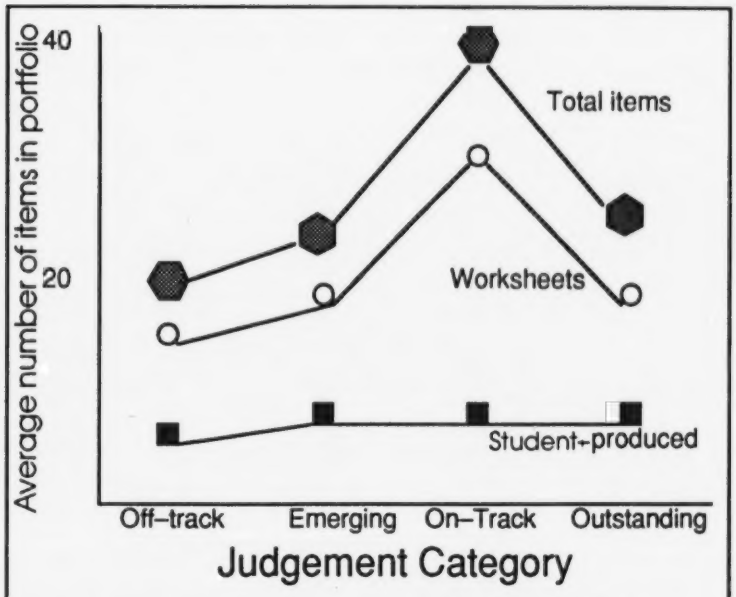

Figure 2. Average number of items found in 42 math portfolios as a function of type of item and judgment category

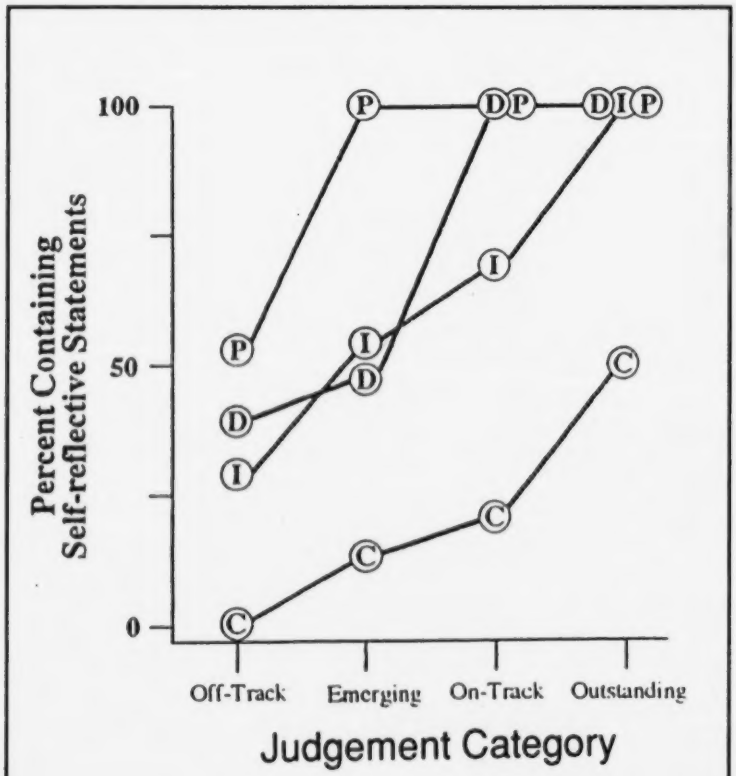

Figure 3 Evidence of self–reflection found in 42 math portfolios as a function of type of self–reflection (*D* Documentation, *C* Comparison, *I* Integration, or *P* Presentation) by judgment category.

fairly constant regardless of quality of portfolio. They did, however, constitute a much larger proportion of the content of the outstanding portfolios.

Finally, we were interested in looking at the role self–reflection played in these portfolios. We asked the team to identify examples of the four categories of self–reflection. We then counted the number of portfolios in each category that contained each of the four types of self–reflection. We converted this to percentages and graphed the results (Figure 3). One thing we discovered is that self–reflective statements are relatively common in all portfolios, even the Off–track ones. We also noted an increase in self–reflection of all four types as the judgment of quality increased.

Generally, the data from the pilot study are cause for cautious optimism. There is evidence that judges can agree with one another when they apply the judgment categories. Moreover, data generated during the pilot study conform to our theoretical assumptions about processes that occur when students assemble quality portfolios.

Conclusion

Current pilot studies using portfolios to assess learning in the aggregate look at portfolios only as a source of examples of student work to be assessed. But assessment drives instruction: Resnick and Resnick (1992) have said that "you get what you assess." By

assessing portfolios as containers of work samples, the evaluation community may be sending the wrong message to the educational community. The portfolio is not just a place to store work samples — it is a self–contained learning environment with valid outcomes of its own. These outcomes must be included in any evaluation using portfolios if the study is to be considered complete. This article presents a rubric designed to infer process and judge the quality of portfolios themselves. The rubric is built on assumptions about the cognitive processes that occur when students build portfolios.

The article reviews a pilot study in which the rubric was used to judge a sample of 42 second grade math portfolios. The results indicate that judges may be able to use this rubric to produce consistent judgments of portfolio quality. The pilot study also includes some preliminary data that relates ratings of the portfolios to other variables of interest.

References

Bereiter, C. and Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Earlbaum.

Bruner, J. (1963). *The process of education*. Cambridge, MA: Harvard.

Eliot Eisner (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. New York: Macmillan.

Frazier, D. L. & Paulson, F. L. (May, 1992). How portfolios motivate reluctant writers *Educational Leadership*, 49(8), 62–65.

Gentile, C. (1992). *Exploring new methods for collecting students' school–based writing*. Washington, DC: National Center for Educational Statistics.

Gearhart, M., Herman, J. L., Baker, E. L., and Whittaker, A. K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment* (CSE Technical Report 337). Los Angeles: University of California, UCLA Graduate School of Education, National center for the Study of Evaluation, Standards, and Student Testing.

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Harman, S. (1992, November). Snow White and the seven warnings: Treats to authentic assessment. *The Reading Teacher, 46(3)*, 250–252.

Kauffman, G. and Short, K. (1993. Self–evaluation portfolios: A device to empower learners. In L. K. Rhodes & N. Shanklin, *Windows into literacy: Assessing learners K–8* (pp. 443–448): Portsmouth NH, Heinemann.

Moss, P. A. (1994, March). Can there be validity without reliability? *Educational Researcher. 23(2)*, 5–12.

Paulson, F. L. & Paulson, Pearl R. (1990, August) How do portfolios measure up? A cognitive model for assessing portfolios. Paper presented at a conference Aggregating Portfolio Data sponsored by the Northwest Evaluation Association, Union, WA. (ERIC Document Reproduction Service ED 334 250, ED 324 329).

Paulson, F. L. & Paulson, Pearl R. (1991, April). The ins and outs of using portfolios to assess performance (Revised). Paper read in a symposium Measurement Issues in Performance Assessment, presented at the National Council on Measurement in Education, Chicago. (ERIC Document Reproduction Service No. ED 334 250).

Paulson, F. L., & Paulson, P. R. (1992–a). *Does this title remind you of Agatha Christie? On assessing self–reflection in portfolios* (Draft). Manuscript available from the authors.

Paulson, F. L. & Paulson, P. R. (1992–b, Fall). The varieties of self-reflection. *Portfolio News, 4(1)*, 1,10,11,13,14.

Paulson, F. L. & Paulson, P. R. (in press). Four varieties of self-reflection. In Yancey, K. (Ed.). *Portfolios: Voices from the field*. Urbana, IL: National Council of Teachers of English.

Paulson, F. L., Paulson, Pearl R., & Meyer, Carol A. (1991, February). What makes a portfolio a portfolio? *Educational Leadership, 48(5)*, 60–63.

Paulson, P. R. & Paulson, F. L. (1991). Portfolios: Stories of knowing. In P. H. Dreyer (Ed), Claremont reading conference 55th yearbook 1991. *Knowing: The power of stories* (pp. 294–303). Claremont, CA: Center for Developmental Studies of The Claremont Graduate School.

Paulson, Pearl R., Paulson, F. L., & Frazier, D. M. (in press). Sarah's portfolio. In Yancey, K. (Ed.). *Portfolios: Voices from the field.* Urbana, IL: National Council of Teachers of English.

Resnick, L. B. & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds), *Changing assessments: Alternative views of aptitude, achievement and instruction*. Boston: Kluwer.

Rief, L. (1990, March). Finding the value in evaluation: Self–assessment in a middle school classroom. *Educational Leadership, 47(6),* 24–29.

Schulman, L. (1992). Portfolios in teacher education: What have we learned. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Stake, Robert (1967). The countenance of educational evaluation. *Teachers College Record, 68 (7),* 523–540.

Vermont Department of Education (1991–a). *Looking beyond the answer: A report on Vermont's mathematics portfolio assessment program*. Montpelier, VT: Author, 120 State Street.

Vermont Department of Education (1991–b). *This is my best: A report on Vermont's writing assessment program*. Montpelier, VT: Author, 120 State St.

Vigotski, L. S. (1962). *Thought and language.* Cambridge, MA: MIT.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed), *Review of Research in Education*, Vol. 17, 1991 (pp. 31-74). Washington: AERA.